DOCUMENT RESUME

ED 402 328 TM 025 830

AUTHOR Bond, Linda A.
TITLE Challenges in the Development of State Assessment Programs that Support Educational Reform.
INSTITUTION North Central Regional Educational Lab., Oak Brook, IL.
PUB DATE May 95
CONTRACT RP91002007
NOTE 19p.
AVAILABLE FROM North Central Regional Educational Laboratory, 1900 Spring Road, Suite 300, Oak Brook, IL 60521-1480.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Accountability; *Educational Change; Elementary Secondary Education; Program Development; Standards; *State Programs; Technical Assistance; *Test Construction; Testing Problems; *Testing Programs; Test Use
IDENTIFIERS Educational Indicators; *Reform Efforts; Stakeholders

ABSTRACT

This paper addresses the educational, technical, legal, and practical challenges states must confront as they consider the content of the assessment, its technical quality, the capacity of educators and the public to use the results of the assessment, the benefits and additional complications of performance testing, and the overall tension betwen the push for uniform standards and local control. The following challenges are explored: (1) different assessment purposes, whether as measuring tool, gatekeeping assessment, part of an indicator system or comprehensive system, or for external testing requirements; (2) technical requirements for quality assessment; (3) improving the capacity of educators and educational stakeholders to use assessment well; (4) the special policy considerations for the use of new testing technologies; and (5) changes in the management of education. As states continue to struggle with the challenge of creating assessment that keeps in step with reform, they are attempting to find a balance between the need for universally accepted standards and the need to allow local educators enough flexibility to meet the needs of their individual students. They also struggle to balance taxpayers' needs for accountability with the need for the state to provide support and technical assistance to schools. (Contains 33 references.) (SLD)

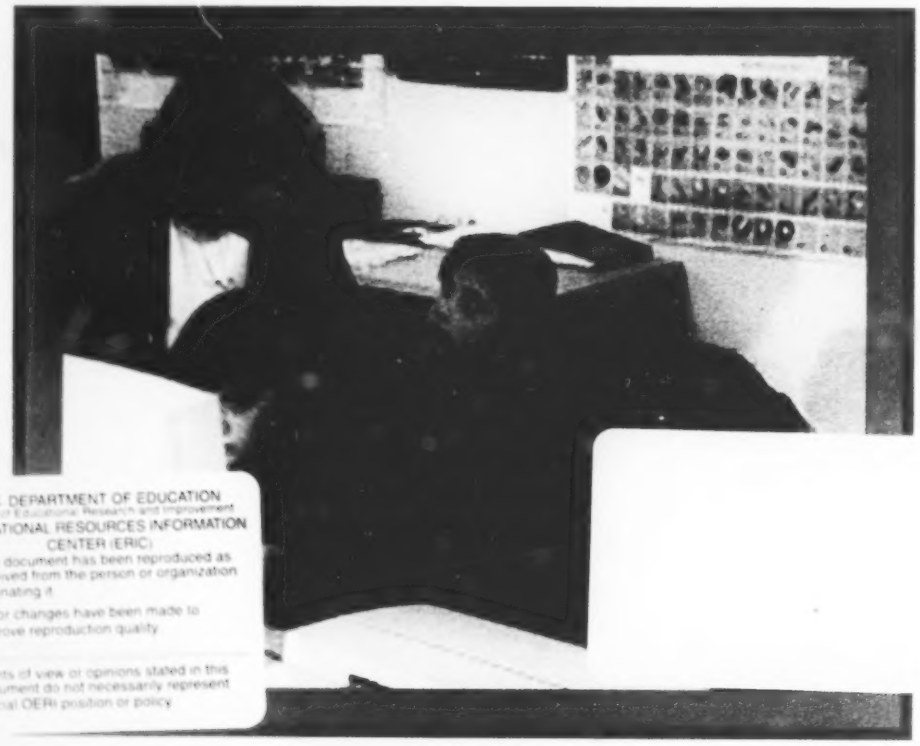
****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 402 328

Challenges in the Development of State Assessment Programs that Support Educational Reform

May 1995

NORTH CENTRAL REGIONAL EDUCATIONAL LABORATORY



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

by Linda A. Bond, Ph.D.
Director of Assessment, Regional Policy Information Center
North Central Regional Educational Laboratory
1900 Spring Road
Oak Brook, Illinois 60521

NCREL

BEST COPY AVAILABLE

TM 025830

North Central Regional Educational Laboratory
1900 Spring Road, Suite 300
Oak Brook, IL 60521
(708)571-4700, Fax: (708) 571-4716

Jeri Nowakowski:	Executive Director
Deanna H. Durrett:	Director, RPIC
Lawrence B. Friedman:	Associate Director, RPIC
Linda Ann Bond:	Director of Assessment, RPIC
Lenaya Raack:	Editor

NCREL is one of ten federally supported educational laboratories in the country. It works with education professionals in a seven-state region to support restructuring to promote learning for all students—especially students most at risk of academic failure in rural and other schools.

The Regional Policy Information Center (RPIC) connects research and policy by providing federal, state, and local policymakers with research-based information on such topics as educational governance, teacher education, and student assessment policy.

© 1995 North Central Regional Educational Laboratory

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number RP91002007. The content of this publication does not necessarily reflect the views of OERI, the Department of Education, or any other agency of the U.S. Government.

Challenges in the Development of State Assessment Programs that Support Educational Reform

Policy Challenges for Assessment

As states struggle to develop an assessment system that attends to their needs for information, they are confronted with challenges and trade-offs. In this paper, the policy challenges that are faced by states as they plan and implement state assessment programs are discussed. This paper addresses the educational, technical, legal, and practical challenges states must confront as they consider the content of the assessment, the technical quality of the assessment, the capacity of educators and the public to use the results of the assessment, the benefits and additional complications that are encountered when performance testing is considered, and the overall tension that exists between the push for uniform standards and local control.[1]

Challenge 1: Different Assessment Purposes

Depending upon the uses made of the results of assessment, the considerations that must be kept in mind by policymakers vary. Purpose is everything, but states described difficulties arising from statutes and rules expecting a single instrument to do too many things. They also expressed concern that assessment instruments that were designed for one purpose were being asked to serve another. Most states are finding that the secret is to determine the purpose(s) of the test and select or design an assessment, or combination of assessments, that are valid for each purpose.[2]

- Measuring Tool or Instrument of Reform

  Chief among the purposes for which states use assessment are two often competing purposes: use as an indicator of educational health and use as an instrument of reform. When assessment is used as an instrument for reform, designed to intentionally cause teachers and students to do something differently in response to the assessment, it loses some of its value as an indicator of educational health. Is the assessment measuring student performance, or is it measuring teachers' preparation of students to perform well on the instrument? In our survey of state assessment directors, the two purposes that states considered most important for their assessment programs were accountability—where the issues of comparability and technical quality are most pronounced—and instructional improvement—where having the assessment match the curriculum of interest is most important. It seems that many states that use assessment for both accountability and instructional reform

purposes find that strengthening the assessment to serve one purpose weakens its utility for the second (Bond & Cohen, 1991; Corbett & Wilson, 1991; Council of Chief States School Officers, et al., 1995; Koretz, Stecher, & Deibert, 1992; Koretz, Stecher, Klein, McCaffrey, & Deibert, 1993b; O'Sullivan, 1991).[3]

- Tests as Gatekeepers of Educational Opportunity

  Those states that have statewide exams, and most do, share a concern that the results of the assessment not be used to deny a student entrance into a higher level course of study. Their concern, and the concern shared by many who believe there are problems with tests, old and new, is that the results are sometimes used to control educational opportunity (Lewis, 1992). Students who may need the most enriching curricula are oftentimes tracked into dead-end remedial classes from which they never emerge. Future educational opportunities are dependent upon previous educational opportunities. Most states have little say over how the results are used in the schools, but many have established policies advising against such tracking.

- Assessment as Part of an Indicator System

  Another issue with which states are struggling is the role testing and assessment plays in an indicator system of educational health. Such a system would need to include many other important indicators, such as the quality of the educational institution's policies and practices (especially those affecting the opportunities and working conditions of teachers and students), the readiness of students for school, the societal support for learning received by the school, the educational and economic support for the school, and the equity of educational opportunity for students (National Center for Educational Statistics, 1991).

- Comprehensive Assessment System

  In searching for ways to improve the match between new educational goals and standards, new curriculum, and state assessment, states are realizing that a comprehensive system of assessments, rather than a single test, can help them address the need to improve the content coverage of state assessment, its utility for a variety of purposes ranging from student certification to instructional modeling, and its match to and support for educational reform (Roeber, 1992).

  It may be that the only solution to the challenge of meeting multiple purposes is to have different but coordinated assessments for different purposes. The challenge will be to see that all the various components of the assessment "system" fit together.

- External Testing Requirements

  Further complicating the development of a comprehensive state assessment system is the need to meet reporting requirements for federally funded educational programs such as Chapter 1, a federal compensatory education program. This

program once required norm-referenced tests for evaluative purposes and many states included a norm-referenced test within their assessment programs. Although the law has been changed (Council of Chief State School Officers, 1994) to allow states to use state assessment programs for evaluative purposes, states are still faced with having an assessment that will allow schools receiving Title 1 funds (renamed from Chapter 1) to demonstrate student academic growth over the course of a year. Although the new law eliminates the norm-referenced testing mandate, states are still faced with needing a state-level exam that will meet the purpose of Title I evaluation. For states with limited funding or with a goal of reducing testing time, their state assessment system will have to include an assessment component that will meet accountability standards of technical quality.

Challenge 2: Technical Requirements for Quality Assessment

No matter what assessment or combination of assessments is used by a state, the technical quality of those assessments is very important. Once again, the challenges will be divided into those that may be considered: educational, technical, legal, and practical.[4]

- Educational

With all the activity in the states around the selection of "learner outcomes" or "essential skills," it is apparent that the content of the assessment or system of assessments is one of the most important and most debated decisions about any state assessment. There is legitimate concern among the states that what is not assessed will not be taught. In the 1980s, many states were using competency tests to identify children in need of extra assistance, and test content was based on minimal standards. New "world-class" standards, the focus for the 1990s, is causing states to ensure that the sample covered by the assessment is *not* a minimal set of objectives nor a "lowest common denominator" of what schools already teach. It is feasible and defensible to allow the test to lead the schools to some extent, but the content chosen must be that which can be measured adequately and that students can learn and teachers teach.

- Technical Issues

All states are struggling with technical issues, and many research studies are being conducted to solve them. The two classic concerns are reliability—knowing that the results of assessments are accurate and stable—and validity—knowing that what we say the assessment is telling us is what the results actually mean. The construction, administration, scoring, and reporting of assessment results are all activities that are governed by the *Standards for Educational and Psychological Testing* (American Educational Research Association, et al., 1985).

(1) Reliability—Differences in test scores or other assessment results should

be related to the differences in the knowledge and skills of the test takers, not due to irrelevant factors such as scoring errors and familiarity with the test content. Sources of unreliability with which states are struggling include: rater bias (the individual rater's biases go into the rating); administration differences (two students taking the same test under very different conditions may not have the same opportunity to demonstrate what they know); and lack of comparability due to different choices of assessment content (in situations where there is a choice about the content of an assessment; unless the choice is based on clear criteria or agreed-upon standards, the resulting assessments may measure very different things). Several of these sources of unreliability can be addressed with professional development opportunities for those who will be administering and scoring the assessments. This is something many states are trying to do.

(2) Validity—The need is to provide evidence to support claims that the test is measuring what it purports to measure, and that the inferences being made from the test scores are justified. There are two major issues that must be addressed by states. First of all, does the assessment or test sufficiently sample from the content being tested to justify its name—for example, reading test, writing test, literacy test—and does the content match the intended outcomes or goals of instruction? Second, are the claims being made about what the test results mean justified? If the test results are used for a specific purpose—entry into a special program, grade level promotion—there must be research evidence that the assessment is accurately identifying students who will or will not succeed.

(3) The Need for Longitudinal Data—Schools and states compare their students' performance over the years to notice any trends in improvement or decline. In order to do this, the assessment must be linked in some way from year to year so that results can be compared. If totally different content was used in year two as opposed to year one, growth or decline would be impossible to gauge. This makes the ability to link performance on one assessment with that on a newer assessment important. Having uniform educational goals against which to judge year-to-year progress would be another way to ensure comparability, but at least a portion of the assessment would have to remain constant over time (or otherwise be equatable). Phasing in and phasing out changes in assessment, and linking scores or performance ratings to an imbedded portion of the assessment from year to year, are two ways states are trying to address this concern (Bond, Friedman, & van der Ploeg, 1993).

- Legal Issues

Anytime a state uses a test for accountability purposes, particularly when those purposes include awarding a high school diploma or certificate of mastery, that test is subject to challenges in court. The courts usually depend upon the *Standards for Educational and Psychological Test-*

ing (American Educational Research Association, et al., 1985) which were written prior to the emergence of so much interest in performance assessments. These standards will apply no matter what kind of assessment is used for these purposes, and evidence of technical quality and use of the assessment for only validated purposes will continue to determine whether an assessment system is upheld in court (Phillips, 1993). Most researchers urge that the necessary studies be conducted on newer assessments to ensure their validity for accountability purposes *before* they are used.

- Practical Considerations

Nearly every interviewee mentioned the need for more time, money, and staff in order to do all that needs to be done to design, develop, and implement an educationally and technically sound assessment program. These resources are needed to involve all of the relevant groups and conduct all of the consensus building and public awareness efforts, technical quality studies and field tests, professional development activities, and program management actions that are necessary to a quality program that is accepted by those most interested in its outcome.

Hidden costs are sometimes not considered in legislation, and state education agencies struggle to keep costs manageable. The costs of conducting the research that is necessary to design, implement, and score a performance exam are considerably more expensive than the cost of buying an off-the-shelf standardized test (Office of Technology Assessment, [OTA]1992). However, the differences in the benefits of the two in terms of enhancement of instruction and professional development opportunities would have to be factored in to get a fair estimate. There are ways to balance the two, for example, using a multiple choice exam to measure those things that can be measured with this approach and using more appropriate kinds of assessment for those outcomes that cannot be measured in this way. Another cost-cutting strategy is evident in interstate collaboratives, such as the CCSSO State Collaborative on Assessment and Student Standards, and the New Standards Project (Learning Research and Development Center, et al., 1992) in which states and school districts join resources with others to share the costs of research and development for new assessments.

Challenge 3: Improving the Capacity of Educators and Educational Stakeholders to Use Assessment Well

- Preservice, Staff Development, Technical Assistance

In nearly every interview with testing directors or directors of educational reform, the single most important benefit of and challenge to state assessment was professional development for teachers and technical assistance to schools. Understanding how to administer, score, and interpret the results of assessment accurately was chief among the concerns

of test directors, particularly those who were working with nontraditional assessment. The reliability of these assessment results are dependent upon the amount and quality of the professional development and follow-up technical assistance received by those doing the scoring. No state felt like they had enough resources, in time and money, to do as much of this as they would like.

While most of the states reported that they were providing some professional development to practicing teachers and administrators, they expressed concern that very little assessment training and instructional reform was taking place at the preservice level. One way that states believe the goals of reform can be promoted is to work with preservice teachers. The governing boards of K-12 public education and higher education, which in many cases are separate government agencies, are working together to improve the linkages between teacher and administrator education and the reform goals of the state.

- Public Awareness

Another major challenge to state education agency personnel is the need to help the public, including legislators, office holders, and the business community, understand what tests can and cannot do, and what their messages are. Too often, too much faith is put into a single test score or statistic, and that one number is expected to tell the public, the school, the teachers, and the students everything about an individual child, school, or school district. States worry that their assessment programs get burdened with so many responsibilities that each time they try to adjust to meet a new responsibility, the usefulness of the assessment for another purpose is diminished.[5] Several states suggested that if the legislative focus is sharpened, some of the overuse and misuse of assessment can be avoided.

Challenge 4: Special Policy Considerations for the Use of New Testing Technologies

The disenchantment with traditional, multiple-choice tests has led many states, districts, and schools to design new testing technologies. These carry with them a host of new issues.

- Nontraditional Assessment

While these new testing technologies may more closely align with new standards than with traditional assessments, their newness complicates the assessment debate (Mehrens, 1992b). These nontraditional tests include essay exams, which have been around for years but are now being refined to yield more precise information about preferred essay characteristics; performance assessment, where students perform the desired behavior and that performance is rated (examples include laboratory experiments, classroom projects, and speeches); and computer-adapted testing, where students take a test and have the item difficulty and test content of the rest of the test matched to their

readiness for the next level of content. The use of performance assessments and portfolios (carefully selected samples of student work) in high-stakes assessment is troublesome, however, partly because agreement about quality control criteria has not been reached and these assessments often do not meet traditional quality criteria. However, researchers are working to develop these criteria and to define them in ways that will make them clearly understood and applicable (Linn, Baker, & Dunbar, 1991).

- Educational Issues

Because the tasks are more authentic (relevant to the kinds of tasks students do outside and inside of school), and because the assessment requires students to produce (rather than passively select) a solution, the assessment reinforces improved approaches to instruction (Mitchell, 1992; Wiggins, 1989). Clearly, this is the good news. Advocates recognize the power that state and national tests can have over what teachers teach, and hope to use this form of assessment to promote good instruction (Mitchell, 1992; Wiggins, 1989). However, other researchers are pointing out that the content of performance tests can also be taught to inappropriately in high stakes situations. They also say that staff development in new instructional approaches will be necessary if a positive impact on classrooms is to be achieved (Koretz, Stecher, Klein, McCaffrey, & Deibert (1993); Stiggins, 1990).

- Technical Issues

Even those who are involved in the development of these new technologies urge caution in trying to use these assessments for multiple purposes before they are ready (Aschbacher, 1991; Dunbar, Koretz, & Hoover, 1991; Koretz, Klein, McCaffrey, & Stecher, 1993; Mehrens, 1992b; Quellmalz, 1991; Reidy, Yen, Gabrys, Hill, & Haertel, 1993). However, many believe that the educational benefits make continued development worth it.[6] Instead, they recommend wide-scale research and refinement at the state and national level and the use of these assessments at the school and classroom level, at least to start. "Simply because the measures are derived from actual performance or relatively high-fidelity simulations of performance, it is too often assumed that they are more valid than multiple-choice tests" (Linn, Baker, & Dunbar, 1991, p. 16).

- Legal Issues

Legally, students cannot be held accountable for performance on an assessment that contains material in which they have not received instruction. In addition, the technical quality of the assessment must be proven in court if the test is challenged. Many urge caution, and urge that resources be devoted to the continuing research to improve the utility of performance assessment as a measurement device (Madaus, 1991; Mehrens, 1993a. Shepard, 1991).

- Practical Issues

  In addition to massive experimentation and field study, performance assessment will also require professional development in order to help teachers shift their instructional approaches, and understand how to use and interpret the assessment results. Without sufficient opportunities to learn the new instructional and assessment strategies, teachers may end up "teaching to the authentic test" in ways that will not result in improved learning for students (Madaus, 1991; Shepard, 1991; OTA, 1992). Scoring is also quite expensive for performance assessments, although advocates remind us that scoring is a professional development opportunity for teachers who rate the quality of students' performances.

Challenge 5: Changes in the Management of Education

- Effect on Teacher Flexibility

  One of the unintended consequences of high-stakes assessment has been to limit the flexibility of teachers' decisionmaking. A single test score is sometimes considered more important than a year's worth of teacher judgment when the two are in conflict. When the goal of a state standards and assessment program is to improve the ability of educators to meet the needs of their students, teachers are fairly comfortable with the program (Darling-Hammond & Wise, 1985; Stake & Theobald, 1991). In some states, teachers were fairly comfortable with the state program because they thought uniform standards would refocus instruction. Their feelings have changed because, in some cases, the tests have been used to evaluate teachers, to label schools "inferior," to offer money for performance, and, in general, to criticize teachers and schools (Corbett & Wilson, 1991). Excessive emphasis is placed on the test when it is used for such high-stakes purposes, and teachers end up focusing instruction primarily on test content (Smith, 1989).

  Programs like the portfolio program in Vermont, which is voluntary and calls for tremendous teacher involvement, is well received by teachers except for concerns about the amount of time involved. Still, almost every district in Vermont participated and expanded the portfolio beyond the grade levels included in the state program (Koretz, McCaffrey, Klein, Bell, & Stecher, 1992). Their difficulty has been in the lack of uniformly selected portfolio content and inadequate training in scoring, problems they are working to overcome. Several states are moving toward a system of state standards and assessments, with more flexibility and involvement of teachers in the design and implementation (for example, California, Kentucky, and Vermont).

- The Balance Between Uniform Standards and Local Control

  In the 1980s, standards for student performance were determined state by state and were sometimes simply cut scores on state tests. Most students passed those tests, even though there was a growing

awareness among educators and the public that students did not possess the skills they would need to survive in a highly technical, globally competitive society. Efforts are in place across the majority of the states to adopt new, realistic standards for success in the twenty-first century. Because ours is such a mobile society with students moving from place to place, it is difficult to imagine every school in the country teaching different material and expecting different levels of performance. On the other hand, with as much diversity as we have, it is also hard to imagine every school teaching exactly the same material at exactly the same pace with exactly the same expectations for all students. Some argue that parents and students need "external standards against which they can measure the performance of their children and their children's schools" (Office of Educational Research and Improvement, 1992).

Most of the states in our study reported that the balance between uniform standards and the rights of a school or school district to set curriculum for its students is sometimes difficult to achieve. Most have adopted a set of "essential skills" believing that most of the schools and those in the public will agree that there is a core set of skills that all students should possess. Schools and districts are then free to supplement this core with locally selected standards. In addition to state standard-setting efforts, the National Council on Education Standards and Testing (National Council on Educational Standards and Testing, 1992) called for the development of a voluntary nationwide system of assessments that is linked to national standards for each of the five core subjects of English, mathematics, science, geography, and history. States struggle to tie local assessments to a common standard, and these difficulties will only be exacerbated at the national level. How these national standards, state standards, and local standards will be linked so that students and schools end up with a coherent set of goals will be a major challenge at all levels.

Summary of Challenges

States face similar challenges regardless of the type of assessment system they implement. Differences in the choices states make are influenced by:

(1) Differences in the purposes for the assessment and competition among purposes, particularly the competition between the two most common purposes for state assessment—accountability and instructional support.

(2) Differences in the state's ability to deal with the educational, technical, legal, and practical issues involved in the implementation of any student assessment program.

(3) Differences in the capacity of teachers, administrators, policymakers, and the public to understand and use assessment appropriately.

(4) Differences in the state's ability to deal with the increasingly complex educational, technical, legal, and practical

issues as they relate to newer testing technologies.

(5) Differences in the tradition of local versus centralized control of education in the state and the need for state assessment to support uniform "world class" standards.

Clearly defining the purpose, or purposes, of a state's assessment program is an important first step in designing a system that will best meet the needs of the state. Most states want assessment information for a variety of purposes, including accountability, information sharing/monitoring, and instructional improvement. A single assessment is sometimes expected to yield all of this information, but it cannot. Whatever decisions about assessment are made by states, tradeoffs are inevitable. Many are finding that a collection of assessments for accountability, monitoring, and instructional improvement appears to be an alternative worth considering. Ensuring the fit among the various components of the assessment system, and keeping the volume of assessment from getting out of hand, will likely be the next challenges states face as they restructure their student assessment programs.

To date, these attempts to develop a comprehensive assessment system have been thwarted by technical, legal, and practical restrictions on what states can do given current research and resources. Technical requirements for an educationally meaningful and legally defensible assessment program entail developmental costs and time, both of which are in short supply in state education agencies. Funding to provide the research and professional development that are needed to change to a different system simply aren't available to many states. For this reason, many states are choosing to supplement rather than supplant their existing assessment program. Nationally norm-referenced tests are still used widely to provide national comparisons, and criterion-referenced tests (mostly multiple choice) are still the norm for measuring agreed upon student learning objectives. Many states are actively experimenting with the use of performance assessment as a part of their state assessment programs, and a few are aggressively pursuing this as a replacement for traditional assessment. Ongoing research and development is needed, and individual states are going to find it increasingly difficult to find the resources to do this alone. Expansion of collaborations with other states, state agencies, universities, private contractors, and research institutions is likely.

For the many states that are working to improve their student assessment programs, there is a universally understood need to increase the capacity of users (educators and stakeholders) to use and understand assessment. Expecting too much of a single instrument and over-interpretation are common misuses of assessment that occur because of a lack of understanding. The newer assessment strategies will require even greater involvement and understanding on the part of users, and professional development and public awareness campaigns will be needed. With limited resources, this too will be a challenge.

New assessments make all of these challenges for states, particularly those related to assessment purposes and technical requirements, even more complex. Still, the educational demands for improving the match between assessment and instructional goals mean that many states will continue to pursue assessments that enable students to construct their own solutions to problems. Refinement through experimentation and field testing will need to be accompanied by professional development to ensure accurate results and proper interpretations.

As states continue to struggle with the challenge of creating assessment that keeps in step with reform, they are attempting to find a balance between the need for universally accepted "world-class standards" for all students, and the need to allow local educators enough flexibility to meet the needs of their individual students. Similarly, states are searching for an appropriate balance between taxpayers' need for accountability information, and the need for the state to provide support and technical assistance to schools. The ultimate effect of this balancing act on state assessment is yet to be determined, but as uniform standards provide the link between national assessment, state assessment, local district assessment, and classroom assessment, the reality of a comprehensive student assessment program is more likely.

References

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. (1985). *Standards for educational psychological testing*. Washington, DC: American Psychological Association.

Aschbacher, P. (1991). Performance assessment: State activity, interest, and concerns. *Applied measurement in education*. 4(4), 275-288.

Bond, L., Friedman, L., & van der Ploeg, A. (1993). *Surveying the landscape of state educational assessment programs*. Oak Brook, IL: North Central Regional Educational Laboratory.

Bond, L. A., & Cohen, D. A. (1991). *Administrators' perceptions of the early impact of Indiana statewide testing for educational progress*. In R. Stake (Ed.), *Advances in program evaluation: Effects of mandated assessment on teaching*, (Vol. 1B, pp. 75-100). Greenwich, CT: JAI Press, Inc.

Corbett, H. D., & Wilson, B. L. (1991). *Testing, reform and rebellion*. Norwood, NJ: Ablex Publishing Corporation.

Council of Chief State School Officers. (1994). Summary of H.R. 6, The *Improving America's Schools Act of 1994*, Reauthorization of the Elementary and Secondary Education Act of 1965. Washington, DC: Author.

Council of Chief State School Officers and North Central Regional Educational Laboratory. (1994). State student assessment program database. Oak Brook, IL: North Central Regional Educational Laboratory.

Darling-Hammond, L., & Wise, A. E. (1985, January). Beyond standardization: State standards and school improvement. *The Elementary School Journal*, 315-336.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education*, 4(4), 289-308.

Koretz, D., Klein, S., McCaffrey, D., & Stecher, B. (1993a). *Interim report: The reliability of Vermont portfolio scores in the 1992-93 school year* (October 18, 1993). Washington, DC: RAND Institute on Education and Training, and Los Angeles, California: National Center for Research on Evaluation, Standards, and Student Testing.

Koretz, D., McCaffrey, D., Klein, S., Bell, R., & Stecher, B. (1992). *The reliability of scores from the 1992 Vermont portfolio assessment program*. Washington, DC: RAND Institute on Education and Training.

Koretz, D., Stecher, B., & Deibert, E. (1992). *The Vermont portfolio assessment program: Interim report on implementation and impact, 1991-1992 school year.* Washington, DC: RAND Institute on Education and Training.

Koretz, D., Stecher, B., Klein, S., McCaffrey, D., & Deibert, E. (1993). *Can portfolios assess student performance and influence instruction?* The 1991-92 Vermont Experience (December, 1993). Washington, DC: RAND Institute on Education and Training, and Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Linn, R.L., Baker, E.L., Dunbar, S.B. (1991). Complex, Performance-Based Assessments: Expectations and Validation Criteria. *Educational Researcher, 20(8),* 15-21.

Learning Research and Development Center and the National Center on Education and the Economy. (1991). *The New Standards Project.* Pamphlet available from Lauren Resnick, LRDC, University of Pittsburg.

Madaus, G. F. (1991). The effects of important tests on students: Implications for a national examination system. *Phi Delta Kappan,* 226-231.

Mitchell, R. (1992). Testing for learning: From testing to assessment. In *Perspective,* Council for Basic Education, 4(2).

Mehrens, W. A. (1992a). *A policymaker's guide to high school graduation testing.* Oak Brook, IL: North Central Regional Educational Laboratory.

Mehrens, W. A. (1992b, Spring). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice,* 3-20.

National Center for Educational Statistics. (1991). *Education counts: An indicator system to monitor the nation's educational health.* Washington, DC: Author.

National Council on Education Standards and Testing. (1992). *Raising standards for American education: A report to Congress, the Secretary of Education, the National Education Goals Panel, and the American people.* Washington, DC: Author.

Office of Educational Research and Improvement. (1992). *Parental satisfaction with schools and the need for standards.* Educational Research Report, Washington, DC: U.S. Department of Education. (OR 92-3070).

Office of Technology Assessment. (1992). *Testing in American schools, Asking the right questions* (297-934 QL 3). Washington, DC: Congress of the United States.

O'Sullivan, R. G. (1991). Teachers' perceptions of the effects of testing on classroom practice. In R. Stake (Ed.), *Advances in program evaluation: Effects of mandated assessment on teaching,* (Vol. 1B, 1993, pp. 145-162). Greenwich, CT: JAI Press, Inc.

Phillips. S. E. (1993). *Legal implications of high stakes assessment*: What states should know. Oak Brook, IL: North Central Regional Educational Laboratory.
Quellmatz, E.S. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education*, 4*(4)* 319-331.

Reidy, E., Yen, W., Gabrys, R., Hill, R., & Haertel, E. (1993). *The use of performance assessment in high-stakes environments: Is there sufficient technical quality for high-stakes usage?* A presentation at the CCSSO National Conference on Large Scale Assessment, June 8-10, 1993, Albuquerque, NM.

Roeber, E. (1992). *How should the comprehensive assessment system be designed?: A. Top Down, B. Bottom up, C. Both, D. Neither. Unpublished manuscript available from the author at the Council of Chief State School Officers, Washington, DC.*

Shepard, L. (1989). *Inflated test score gains: Is it old norms or teaching to the test?* Paper presented at the annual meeting of the American Educational Research Association, March, 1989, San Francisco, CA.

Shepard, L. A. (1991, November). Will national tests improve student learning? *Phi Delta Kappan*, 232-238.

Smith, J. (1989) *What you test is what you get!* Presentation at the Indiana Policy Seminar, Planning for the Future of Indiana Statewide Testing for Educational Progress, Indianapolis, IN, June 23, 1989.

Stake, R., & Theobald, P. (1991). Teacher's views of testing's impact on classrooms. In R. Stake (Ed.), *Advances in program evaluation: Effects of mandated assessment on teaching*, (Vol. 1B, pp. 189-201). Greenwich, CT: JAI Press, Inc.

Stiggins, R. J. (1990). The foundation of performance assessment: A strong testing program. In *Policy Briefs*, No. 10 & 11. Oak Brook, IL: North Central Regional Educational Laboratory.

Wiggins, G. (1989, April) Teaching to the test. *Educational Leadership*, 41-47.

Endnotes

[1]Most of the information about states used in this paper can be found in the State Student Assessment Program Database (1992-93; 1993-94; 1994-95). The database provides information about state assessment programs which has been collected by survey from state assessment directors (the Association of State Assessment Programs (ASAP), including detailed information about each component of the state's assessment program; the assessment design, format, and purpose; the use of nontraditional assessment methods; and the state's plans for the future of the program. (Council of Chief State School Officers and North Central Regional Educational Laboratory, 1995).

[2]A cogent, brief discussion of issues surrounding tests serving varoius needs and purposes can be found on pages 10-12 of *Testing in American Scools: Asking the Right Questions* (Office of Technology Assessment, 1992).

[3]For example, some have argued that minimizing the pressure to teach to the test will improve its utility as an accountability measure. Strategies such as keeping the test secure, giving the test to only a sample of the students, and using the assessment early in the year diminish the likelihood of teaching to the test, but also diminish the likelihood that teachers can use the results to improve instruction (Shepard, 1989). If they do not know what is on the test, and if they do not have scores for each of their students, it will be difficult to improve instruction for those students. A balance is needed.

[4]This section relies heavily on *Testing in American Schools: Asking the Right Questions* (Office of Technology Assessment 1992) and a North Central Regional Educational Laboratory report, *A Policymaker's Guide to High School Graduation Testing* (Mehrens, 1992a).

5See the section "Challenge 1: Purposes of Assessment."

[6]States like Vermont are seeing improvements in classrooms across their state and believe that the effort to make these new assessments more reliable is worth the positive consequences for students. Still they caution against using the results for anything more than a state profile until the reliability of scoring is improved (Dunbar, Koretz & Hoover, 1991; Koretz, Stecher, & Deibert, 1992; Koretz, McCaffey, Klein, Bell, & Stecher, 1992)

North Central Regional Educational Laboratory
1900 Spring Road, Suite 300
Oak Brook, IL 60521-1480
(708) 571-4700, Fax (708) 571-4716

BEST COPY AVAILABLE